UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

MW

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/801,968	03/16/2004	Silviu-Petru Cucerzan	MS306752.1/MSFTP585US	2547

27195 7590 10/11/2006

AMIN. TUROCY & CALVIN, LLP
24TH FLOOR, NATIONAL CITY CENTER
1900 EAST NINTH STREET
CLEVELAND, OH 44114

EXAMINER	
TSUI, WILSON W	
ART UNIT	**PAPER NUMBER**
2178	

DATE MAILED: 10/11/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)
	10/801,968	CUCERZAN ET AL.
	Examiner	Art Unit
	Wilson Tsui	2178

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *20 July 2006*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-12,14-16,18-32,34-36,38,41 and 42* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-12, 14-16,18-32,34-36,38,41 and 42* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

1. This action is in response to the amendment filed on: 7/20/2006.

2. Claims 1, 3, 4, 5, 14, 15, 18, 32, 34, 36, 38, 41, and 42 have been amended. Claims 13, 17, 33, 37, 39, and 40 have been cancelled. Claims 1, 32, 38, 41, and 42 are independent claims.

3. The 35 USC 101 rejections for claims 39 and 40 have been withdrawn, since claims 39 and 40 have been cancelled.

4. The 35 USC 112 rejections for claims 3, 4, 40, 41, and 42 have been withdrawn.

5. Each of the claims (claims 17, 25, 31, and 37) that were objected to as being dependent on a rejected base claim for having allowable subject matter, are no longer objected to, in view of new grounds of rejection.

Claim Rejections - 35 USC § 112

The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

6. Claim 14 is rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

Claim 14 recites the limitation "the group" in page 3 of the amended claims. There is insufficient antecedent basis for this limitation in the claim.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

1. Claims 1-4, 6-12, 14, 15, 18, 20-24, 38, and 42 are rejected under 35 U.S.C. 103(a) as being unpatentable over Tang et al (US Patent: 6,636,849 B1, issued: Oct. 21, 2003, filed: Nov. 22, 2000), Shanahan et al (US Application: US 2003/0033288 A1, published: Feb. 13, 2003, filed: Dec. 5, 2001), and Beeferman et al (US Patent: 6,701,309 B1, issued: Mar. 2, 2004, filed: Apr. 21, 2000) in further view of Hitachi (Derwent, published: Feb 16, 2001, Abstract).

With regards to claim 1, Tang et al teaches a system that facilitates spell checking comprising:

- *A component that receives input data containing text* (column 4, lines 55-66: whereas a search string is received)
- *A spell checking component that identifies potentially misspelled strings in the text, and proposes at least one alternate spelling for the string* (column 7, lines 20-30: whereas, the Tang et al's system teaches spell checking potentially misspelled words using a dictionary/lexicon, and returning a suggestion to the user concerning a least one alternate spelling.)

However, Tang et al does not teach creating/using *substrings* of the text, and providing an alternate spelling for the *substring set*, *based on at least one query log; the query log comprising data utilized by users to query a data collection over a time frame, the spell checking component utilizes occurrence and co-occurrence statistics from the*

at least one query log, the substring co-occurrence statistics comprising substring bigram counts with stop-word sequence skipping counts.

Shanahan et al teaches a *spell checking system* (paragraph 0518: whereas, Shanahan et al's system takes text, and identifies text that need spelling corrections). The spell checking system takes *substring data from the* input *text* (Fig 3.1: whereas, text from a document is processed by tokenizing words, and identifying N-Gram of words from the input text after removal of stop words). All words (substrings of the input text) are iteratively are processed and corrected to generate a set of alternate spellings for the input text as shown in Fig. 51. *Furthermore, stop-word-sequence-skipping counts* are implemented, to further refine the spell checking process (paragraph 305: whereas, in expert mode, only entities that occur in referenced documents with a (tracked/logged) frequency below a predefined threshold are annotated (for the purpose of detecting and ignoring/skipping stop words (for which the stop words have a count above a predefined threshold)).

It would have been obvious to one of the ordinary skill in the art at the time of the invention to have modified Tang et al's spell checking component to further take and process and provide alternate spellings for substrings of input text (through various techniques, including logged/tracked stop word skipping counts) as taught by Shanahan et al. The combination of Tang et al and Shanahan et al would have allowed Tang et al's system to have "identified errors in a document, by formulating a query using identified errors in document content, identifying a set of entities in the database of entities that satisfies the query; correcting the document content using the identified set

of entities, and updating the information space with the corrected document content" (paragraph 0015).

However, although Tang et al and Shanahan et al teach the implementation of a spell checking component for alternate spelling of a substring set, through various techniques including the use of tracked/logged data (stop-word-skipping-counts), as explained above, they do not expressly teach the alternate spelling of a substring set is *based on at least one query* log; *the query* log *comprising data utilized by users to query a data collection over a time frame, the spell checking component utilizes occurrence and co-occurrence statistics from the at least one query* log, *the substring co-occurrence statistics comprising substring bigram counts.*

Beeferman et al teaches an alternate spelling of a string query (column 1, lines 51-54: whereas, a system for query refinement includes suggesting an alternate spelling or a corrected spelling for a query) is *based on data stored in* a *query log file* (columns 10 and 11, lines 52-64 and 1-10 respectively: whereas, based on heuristic data from query log data, it is determined if a suggested spelling is appropriate), *the query log comprising data utilized by users to query a data collection over a time frame* (Table 2, column 9, lines 45-67, and column 10, lines 1-6: whereas a query log holds data about the number of occurrences for each particular query/string has been submitted by a particular class of users in searches, over a period of time), *and substring occurrence and co-occurrence statistics from the at least one query log*: in Table 2 (whereas, the query log comprises the number of occurrences for each particular query (query substring pair) requested. Furthermore, the number of occurrences are shown to be

greater than one, and thus, co-occurrence counts for each of the particular queries are also recorded.

It would have been obvious to one of the ordinary skill in the art at the time of the invention to have modified Tang et al, and Shanahan et al's spell checking component such that spell checks using substrings (and logged/tracked word-skipping-count data), to further include the logged tracked data with the heuristic query log data, that is taught by Beeferman et al. The combination of Tang et al, Shanahan et al, and Beeferman et al would have allowed Tang et al's system to have been able to have implemented a spell checking system that would have "refined a presentation of an alternative query to a first query based on a searcher's tendency to utilize information" (column 2, lines 27-30), and to have also " collected related queries that have a likelihood of being submitted by a class of searcher" (column 2, lines 24-26).

However, although Tang et al, Shanahan et al, and Beeferman teach the *query log statistics* data as explained above, they do not expressly teach the query log data *comprises substring bigram counts.*

Hitachi teaches the statistics comprising *substring bigram counts;* a substring bigram comprising a pair of substrings in a text (Abstract: whereas, a collecting unit collects substrings / bigram strings from a document, and a counter counts the occurrence(s) for the pair of bi-grams).

It would have been obvious to one of the ordinary skill in the art at the time of the invention to have modified Tang et al, Shanahan et al, and Beeferman et al's substring co-occurrence statistics such that they also include bigram counts from a pair of

substrings in a text/document, as taught by Hitachi. The combination of Tang et al, Shanahan et al, Beeferman et al, and Hitachi, would have allowed Tang et al's spell checking component to have been able to evaluate each pair of bigrams "in order of degree of importance" (Hitachi, Abstract).

With regards to claim 2, which depends on claim 1, for *a spell checking component further utilizes user-dependent information in proposing at least one alternative spelling*, is similarly taught by Tang et al, Shanahan et al, Beeferman et al, and Hitachi, in claim 1, and is rejected under the same rationale.

With regards to claim 3, which depends on claim 1, Tang et al teaches *the alternative spelling for the substring set is further based on at least one trusted lexicon* with content (column 7, lines 23-29: whereas, a dictionary which comprises the correct spelling of words, is used as a basis for providing an alternative spelling).

With regards to claim 4, which depends on claim 3, Tang et al teaches the *spell checking component*, in claim 1, and is rejected under the same rationale. However Tang et al does not teach the spell checking component *further employs a list of stop words*.

Shanahan et al teaches a *list of stop words* with content (paragraph 0365: whereas, a set/list of stop words are used to normalize input text data for contextual classification).

It would have been obvious to one of the ordinary skill in the art at the time of the invention to have modified Tang et al's spell checker such that the input text data will be normalize by removing stop words as taught by Shanahan et al. The combination would

have allowed Tang et al's spell checker to have been able to remove stop words "that do not improve the quality of classification" (paragraph 0365).

With regards to claim 6, which depends on claim 4, Tang et al teaches a *spell checking component*, in claim 1, and is rejected under the same rationale. Furthermore, Tang et al teaches *an iterative process to search a space of alternative spellings* (Fig 6, column 12, lines 1-30: whereas, processing for exact or inexact matches are performed on a search tree start, and iterations or a loops take place until the last level of a search tree is reached (looping occurs from reference numbers 630 to 670 and then back to 630 in Fig 6).

With regards to claim 7, which depends on claim 6, Tang et al teaches a *spell checking component*, in claim 1, and is rejected under the same rationale. Furthermore, Tang et al teaches *at least in part, heuristics to impose restrictions on a search space utilized to determine a proposed alternative spelling* (column 15, lines 60-67, and column 16, lines 1-5: whereas heuristic methods are used to impose restrictions on a search space by calculating a distance score that is used in determining candidates for alternative spellings)).

With regards to claim 8, which depends on claim 7, Tang et al teaches *the heuristics* in claim 7, and is rejected under the same rationale. Furthermore, Tang et al teaches the heuristics *utilize, at least in part, at least one fringe to limit the search space* (column 9, lines 59-67, and column 10, lines 1-5: whereas, several fringes are implemented to limit the search space, such as the probabilistic distance function having to be positive, and a triangle inequality has to be satisfied).

With regards to claim 9, which depends on claim 4, Tang et al, Shanahan et al,, Beeferman et al, and Hitachi similarly teach *the query log comprising a histogram of queries asked over a time frame*, as explained in claim 1, and is rejected under the same rationale.

With regards to claim 10, which depends on claim 9, Tang et al, Shanahan et al, Beeferman et al, and Hitachi teach *the histogram of queries*, as explained in claim 9, and is rejected under the same rationale. Tang et al, Shanahan et al, Beeferman et al, and Hitachi also teach the histogram of queries *relates to a subset/*class *of the users*, as explained in claim 1, and is rejected under the same rationale. Furthermore, the *subset/*class *comprises at least one user* (Beeferman et al, column 5, lines 7-9: whereas a particular class of searchers represents a subset of users with at least one searcher/user).

With regards to claim 11, which depends on claim 9, Tang et al, Shanahan et al, Beeferman et al, and Hitachi teach a *query log*, as explained in claim 1, and is rejected under the same rationale. Furthermore, Beeferman et al teaches the query log *resides on a server computer* (column 5, lines 39-40: whereas the query log is downloaded from a search engine/server computer to the client computer, and thus the query log originally resides on the server computer).

With regards to claim 12, which depends on claim 9, Tang et al, Shanahan et al, Beeferman et al, and Hitachi teach *a query log*, as explained in claim 1, and is rejected under the same rationale. Furthermore, as explained in claim 11, the query log is

downloaded from the server to the client, and thus the *query log resides on the client computer as well.*

With regards to claim 14, which depends on claim 1, Tang et al, Shanahan et al, , Beeferman et al, and Hitachi teaches *a substring comprising at least one selected from the group[W1] consisting of an entry in at least one lexicon,* as explained in claim 3, and is rejected under the same rationale.

With regards to claim 15, which depends on claim 1, Tang et al, Shanahan et al, Beeferman et al, and Hitachi teach *a substring bigram comprising a pair of substrings in a text*, as explained in the rejection for claim 1, and is rejected under the same rationale.

With regards to claim 18, which depends on claim 1, Tang et al, Shanahan et al, Beeferman et al, and Hitachi teach *substring co-occurrence statistics from the query log*, as explained in claim 1, and is rejected under the same rationale. Furthermore, the query information is stored in a single data structure/log by downloading from a server as explained in claim 11, and is rejected under the same rationale.

With regards to claim 20, which depends on claim 18, Tang et al and Shanahan et al teach *a spell checking system handling split substrings* by splitting input text into an N-Gram set of words, as explained in claim 1, and is rejected under the same rationale. Furthermore, Shanahan et al teaches a method for using heuristics to determine word similarity, which does not differ/(*operates in the same manner*), if the input text is an individual string or an N-Word split substring using the searching technique that was explained in claim 6 above.

It would have been obvious to one of the ordinary skill in the art at the time of the invention to have modified Tang et al, Shanahan et al's substring spell checking component, to further utilize the string (individual or substring) independent method for searching a search space, as also taught by Shanahan et al. The combination of Tang et al, Shanahan et al, Beeferman et al, and Hitachi would have allowed Tang et al's spell checking component to have been able to have provided for expanded search results if needed by searching split substrings.

With regards to claim 21, which depends on claim 20, Tang et al, Shanahan et al, Beeferman et al, and Hitachi teach *the spell checking component generates a set of alternative spellings that are substrings in a at least one selected from the group consisting of at least one query log and at least one lexicon*, as explained in claim 1, and is rejected under the same rationale.

With regards to claim 22, which depends on claim 21, Tang et al, Shanahan et al, Beeferman et al, and Hitachi teach *the set of alternative spellings comprising a set of alternative spellings*, as explained in claim 1, and is rejected under the same rationale. Furthermore, Shanahan et al teaches the alternative spellings are determined via an *iterative correction process* (paragraph 0511: whereas, through an iterative correction process, text/string/substring in a document gets replaced/corrected with another substring as an alternative spelling. Furthermore, the iterative correction process halts when all the number of errors corrected at a previous iteration is less than a threshold value, and thus the possible alternative spellings are less appropriate than the current set of alternative spellings).

It would have been obvious to one of the ordinary skill in the art at the time of the invention to have modified Tang et al, Shanahan et al, Beeferman et al, and Hitachi's spell correction component, to have further implemented the generation of alternate spellings in an iterative correction process as also taught by Shanahan et al. The combination of Tang et al, Shanahan et al, Beeferman et al and Hitachi would have allowed Tang et al's system to have repeatedly analyzed input text content until a satisfying correction level has been established.

With regards to claim 23, which depends on claim 22, Tang et al, Shanahan et al, Beeferman et al, and Hitachi teach the *iterative correction process*, comprising *a plurality of iterations that change at least on substring to another substring as an alternative spelling, the iterative correction process halts when all possible alternative spellings are less appropriate than a current set of alternative spellings*, as explained in claim 22, and is rejected under the same rationale.

With regards to claim 24, which depends on claim 23, Tang et al teaches *alternative spellings*, in claim 1, and is rejected under the same rationale. Tang et al also teaches the *appropriateness of alternative spellings are computed based on a probabilistic string distance*, as explained in claim 7, and is rejected under the same rationale. Tang et al however, does not teach the appropriateness of alternative spellings are computed based on a *statistical context model.*

Shanahan et al teaches the appropriateness of alternative spellings are computed based on a *statistical context model* (paragraph 0243: whereas the context of the words surrounding a substring/entity is taken into account, and using ranking

methods, only the highest ranked results are kept as appropriate for an alternative spelling).

It would have been obvious to one of the ordinary skill in the art at the time of the invention to have modified Tang et al's spell correction system for providing alternative spellings, to further include the ability to determine appropriateness for alternative spellings based on not only string distance, but through context – statistical analysis as well. The combination of Tang et al, Shanahan et al, Beeferman et al, and Hitachi would have allowed Tang et al's system to have improved the accuracy of alternative spellings by taking the context of the input text into account when providing alternative results.

With regards to claim 38, Tang et al, Shanahan et al, Beeferman et al, and Hitachi similarly teach a system comprising:

- *Means for receiving input data containing text*, as described in claim 1, and is rejected under the same rationale.
- *Means for identifying a set of potentially misspelled substrings in the text and proposing at least one alternative spelling for the substring set based on at least one query log; the query log comprising data utilized by users to query a data collection over a time frame*, as described in claim 1, and is rejected under the same rationale.
- *The means for identifying a set of potentially misspelled substrings in the text utilizes substring occurrence and co-occurrence statistics from the at least one query log, the substring co-occurrence statistics comprising substring bigram counts with stop-word-sequence-skipping counts; a substring bigram comprising*

a pair of substrings in a text, as similarly explained in the rejection for claim 1, and is rejected under the same rationale.

With regards to claim 42, Tang et al, Shanahan et al, Beeferman et al, and Hitachi similarly teach a device employing the system of claim 1, comprising at least one of a computer, as server, and a handheld electronic device, as explained in claim 1, and is rejected under the same rationale.

2. Claims 5 is rejected under 35 U.S.C. 103(a) as being unpatentable over Tang et al (US Patent: 6,636,849 B1, issued: Oct. 21, 2003, filed: Nov. 22, 2000), Shanahan et al (US Application: US 2003/0033288 A1, published: Feb. 13, 2003, filed: Dec. 5, 2001), and Beeferman et al (US Patent: 6,701,309 B1, issued: Mar. 2, 2004, filed: Apr. 21, 2000) and Hitachi (Derwent, published: Feb 16, 2001, Abstract), in further view of de Hita et al (US Patent: 6,081,774, issued: Jun. 27, 2000, filed: Aug. 22, 1997).

With regards to claim 5, Tang et al and Shanahan et al, teach a list of *stop words*, as explained in claim 4, and is rejected under the same rationale. However, Tang et al and Shanahan et al do not teach *the list of stop words containing high frequency words and function words and their frequent misspellings.*

Hita et al teaches a *list of stop/*skip *words containing high frequency words, function words, and their frequent misspellings*: whereas a stop/skip list is implemented for high frequency words and function words (column 1, lines 43-44), and frequent misspellings (column 2, lines 10-12).

It would have been obvious to one of the ordinary skill in the art at the time of the invention to have modified Tang et al, and Shanahan et al's list of stop words to further

include high frequency words, function words, and their frequent misspellings as taught by Hita et al. The combination of Tang et al, Shanahan et al, Beeferman et al, Hitachi, and Hita et al would have allowed Tang et al's spell checking component to have been able to normalize an input data/string set to focus on words that provide more semantic content.

3. Claim 16 is rejected under 35 U.S.C. 103(a) as being unpatentable over Tang et al (US Patent: 6,636,849 B1, issued: Oct. 21, 2003, filed: Nov. 22, 2000), Shanahan et al (US Application: US 2003/0033288 A1, published: Feb. 13, 2003, filed: Dec. 5, 2001), Beeferman et al (US Patent: 6,701,309 B1, issued: Mar. 2, 2004, filed: Apr. 21, 2000) and Hitachi (Derwent, published: Feb 16, 2001, Abstract), in further view of Herz et al (US Patent: 5,754,939, issued: May 19, 1998, filed: Oct 31, 1995).

With regards to claim 16, which depends on claim 15, Tang et al, Shanahan et al, and Beeferman et al, and Hitachi teach *the substring bigram comprising a pair of substrings in text*, as explained in claim 15, and is rejected under the same rationale. However, Tang et al, Shanahan et al, Beeferman et al, and Hitachi do not expressly teach that the bigrams are adjacent substrings in a text.

Herz et al teaches the bigrams are *adjacent* substrings in a text (column 13, lines 28-30: whereas, text is broken into bigrams, which are 2 adjacent words).

It would have been obvious to one of the ordinary skill in the art at the time of the invention to have modified Tang et al, Shanahan et al, Beeferman et al, and Hitachi's spell checking component to further include the ability to process substring bigrams that are adjacent in a text, as taught by Herz et al. The combination of Tang et al, Shanahan

et al, and Beeferman et al, Hitachi et al, and Herz et al would have allowed Tang et al's spell checking component to have been able to process bigrams that are contextually close to each other.

4. Claims 19, 26, 27, 28, 29, 30-36, and 41 are rejected under 35 U.S.C. 103(a) as being unpatentable over Tang et al (US Patent: 6,636,849 B1, issued: Oct. 21, 2003, filed: Nov. 22, 2000), Shanahan et al (US Application: US 2003/0033288 A1, published: Feb. 13, 2003, filed: Dec. 5, 2001), Beeferman et al (US Patent: 6,701,309 B1, issued: Mar. 2, 2004, filed: Apr. 21, 2000), and Hitachi (Derwent, published: Feb 16, 2001, Abstract), in further view of Srihari et al (ACM, published: January 1983, pages 72-75).

With regards to claim 19, which depends on claim 18, Tang et al teaches a tree *data structure* extracted from a lexicon (column 7, lines 21-23). However, Tang et al does not teach a data structure comprising a *trie*.

Srihari et al teaches a data structure comprising a *trie* (Section 3, P3-1, Figure 2: whereas, a data structure used to represent a lexicon is a trie).

It would have been obvious to one of the ordinary skill in the art at the time of the invention to have modified Tang et al's method for representing a lexicon in the form of a trie, as taught by Srihari et al. The combination of Tang et al, Shanahan et al, Beeferman et al, Hitachi, and Srihari et al, would have allowed Tang et al's spell checking component to have implemented a "data structure that is suitable for determining whether a given string is an initial substring" (Srihari et al, Section 3, P3-2).

With regards to claim 26, which depends on claim 24, Tang et al, Shanahan et al, and Beeferman et al teaches *a set of alternative spellings for a substring is generated,*

in as explained in claim 1, and is rejected under the same rationale. Tang et al also teaches a *searchable string data structure extracted from a trusted lexicon* (column 7, lines 22-24: whereas, a structured tree for a whole dictionary/lexicon is created for searching). Furthermore, Beeferman teaches a *searchable query log data structure* (Table 2, whereas, a flat data structure is used to store occurrence and co-occurrence query data). However, Tang et al, Shanahan et al, Beeferman et al, Hitachi do not expressly teach a searchable *substring* data structure.

Srihari et al teaches a searchable *substring* data structure (Page 72-73, Section 3. Lexical Organization, Fig. 2: whereas, a trie is extracted from a lexicon, for which the trie is used to implement a searchable substring data structure using the Viterbi algorithm).

It would have been obvious to one of the ordinary skill in the art at the time of the invention to have modified Tang et al, Shanahan et al, and Beeferman et al's spell correction system such that trie data structures storing substrings are extracted from a particular source (such as a lexicon or query log), such that the alternative spellings are generated from a trie using a Viterbi algorithm, as taught by Srihari et al. The combination of Tang et al, Shanahan et al, Beeferman et al, Hitachi, and Srihari et al would have allowed Tang et al's spell checking system to have used a data structure that is "efficient for text correction algorithms" (Srihari et al, page 72, Section 3).

With regards to claim 27, which depends on claim 26, Tang et al and Shanahan et al teach the processing of *substrings* from input text, in claim 1, and is rejected under the same rationale. Also Tang et al teaches *the set of alternative strings for each string*

query is restricted to within a probabilistic distance from an input string (column 10, lines 32-42: whereas, alternative spellings for a string are based on several factors, including the probabilistic distance)*; the restriction is imposed within each iteration without limiting the iterative correction process as a whole* (column 10, lines 48-60: whereas, the process is repeatedly extended to multiple search spaces or "grids").

It would have been obvious to one of the ordinary skill in the art at the time of the invention to have used Tang et al, and Shanahan et al's method for processing substrings from input text, and additionally used Tang et al's method for iteratively processing the search space with a substring by applying probabilistic distance calculations. The combination of Tang et al, Shanahan et al, Beeferman et al, Hitachi, and Srihari et al would have allowed Tang et al's system to increase the speed and relevancy of possible alternative spellings for a given input substring.

With regards to claim 28, which depends on claim 27, Tang et al, Shanahan et al, Beeferman et al, and Hitachi teaches *the iterative correction process*, in claim 6, and is rejected under the same rationale. Furthermore, Shanahan et al teaches *an iterative correction process searches for an optimum set of alternative spellings* via *utilization of a statistical context model*: whereas the context of the words surrounding a substring/entity is taken into account, and using ranking methods, only the highest ranked results are kept as appropriate for an alternative spelling (paragraph 0243) and the iterative correction process halts when all the number of errors corrected at a previous iteration is less than a threshold value, and thus the possible alternative spellings are less appropriate than the current set of alternative spellings (paragraph

0511: iterative process stops when the number of errors corrected is less than a threshold (optimal value)). Additionally, as explained in the rejection for claim 1, Beeferman et al also teaches a *statistical context model* comprising substring occurrence and co-occurrence statistics extracted from at least one query log (shown in Table 2 of Beeferman et al).

It would have been obvious to one of the ordinary skill in the art at the time of the invention to have further modified Tang et al, Shanahan et al, Beeferman et al, and Hitachi's iterative correction system to further include the ability to use a statistical context model to determine an optimal set of alternative spellings, as taught by Shanahan et al. The combination of Tang et al, Shanahan et al, Beeferman et al, Hitachi, and Srihari et al, would have allowed Tang et al's spell checking component to have been able to iteratively go though a search space, and choosing alternative spellings based on context of the input sentence/string/substring.

With regards to claim 29, which depends on claim 28, Tang et al, Shanahan et al, Beeferman et al, and Hitachi teach *the statistical context model*, comprising *substring occurrence and co-occurrence statistics extracted from at least one query log*, as similarly explained in the rejection for claim 28,and is rejected under the same rationale

With regards to claim 30, which depends on claim 29, Tang et al, Shanahan et al, Beeferman et al, Hitachi, and Srihari et al teach:

- *A Viterbi search is employed to facilitate in determining the optimum set of alternative spellings*, as explained in claim 26, and is rejected under the same rationale.

- Alternate spellings are determined *according to the context model in each iteration*, as explained in claim 28, and is rejected under the same rationale.

With regards to claim 31, which depends on claim 31, Tang et al, Shanahan et al, Beeferman et al, Hitachi, and Srihari teach *the Viterbi search*, as similarly explained in the rejection for claim 30, and is rejected under the same rationale. However the combination of Tang et al, Shanahan et al, Beeferman et al, Hitachi, and Srihari as explained above, does not teach the Viterbi search *can employ fringes to restrict a search for alternate spellings in an iteration such that for every pair of adjacent substrings, if any of the substrings is in at least one trusted lexicon, then only one of the substrings is allowed to change in that iteration.*

Yet, Srihari teaches search *can employ fringes to restrict a search for alternate spellings in an iteration such that for every pair of adjacent substrings, if any of the substrings is in at least one trusted lexicon, then only one of the substrings is allowed to change in that iteration. The Viterbi search can employ fringes to restrict a search for alternative spellings in an iteration such that every pair of adjacent substrings, if any of the substrings is in a least one trusted lexicon, then only one of the substrings is allowed to change in that iteration* (whereas, in a Viterbi search, each iteration corresponds to calculating a single survivor vector/path for a node/state/iteration (page 75). As shown in the for loop, each iteration/state, for each letter/substring of index 'k', of a string of length m, there is an associated cost, that is calculated for possible change using array 'A' in the iteration (page 77, see code 'procedure select(A,C,S,Z)')

It would have been obvious to one of the ordinary skill in the art at the time of the invention to have modified Tang et al, Shanahan et al, Beeferman et al, and Hitachi's context model, to further include Srihari et al's method for implementing a Viterbi context model, which employs fringes. The combination would have allowed Tang et al's spell checking component to have "considered all alternatives for each of the *m* letters" (Srihari et al, page 75).

With regards to claim 32, Tang et al, Shanahan et al, Beeferman et al, and Hitachi similarly teach a method comprising:

- *Receiving input data containing text*, as explained in claim 1, and is rejected under the same rationale.
- *Identifying a set of potentially misspelled substrings in the text*, as explained in claim 1, and is rejected under the same rationale.
- *Generating a set of alternative spellings that are substrings in at least one selected from the group consisting of at least one query log* and *lexicon*: as similarly explained in the rejection for claim 21, and is rejected under the same rationale.
- *The log comprising data utilized by users to query a data collection over a time frame*, as similarly explained in the rejection for claim 1, and is rejected under the same rationale.
- *The set of alternative spellings comprising a set of alternative spellings determined via an iterative correction process*, as similarly explained in the rejection for claim 22, and is rejected under the same rationale.

- *That includes searching for an optimum set of alternative spellings via utilization of a statistical context model* (Beeferman et al, Table 2, lines 48-64: whereas, using statistical context query log data, a set of alternative spellings are identified)
- *The statistical context model comprising substring occurrence, and co-occurrence statistics extracted from at least one query log*, as similarly explained in the rejection for claim 1, and is rejected under the same rationale.
- *Proposing at least one alternative spelling for the substring set*, as explained in claim 1, and is rejected under the same rationale.

However, Tang et al, Shanahan, et al, Beeferman et al, and Hitachi do not expressly teach *employing a Viterbi search to facilitate in determining the optimum set of alternative spellings according to the context model in each iteration; the Viterbi search can employ fringes to restrict a search for alternative spellings in an iteration such that every pair of adjacent substrings, if any of the substrings is in a least one trusted lexicon, then only of the substrings is allowed to change in that iteration.*

Srihari et al teaches:

- *Employing a Viterbi search to facilitate in determining the optimum set of alternative spellings*, as similarly explained in the rejection for claim 26, and is rejected under the same rationale.
- *... according to the context model in each iteration. The Viterbi search can employ fringes to restrict a search for alternative spellings in an iteration such that every pair of adjacent substrings, if any of the substrings is in a least one*

trusted lexicon, then only one of the substrings is allowed to change in that iteration (whereas, in a Viterbi search, each iteration corresponds to calculating a single survivor vector/path for a node/state/iteration (page 75). As shown in the for loop, each iteration/state, for each letter/substring of index 'k', of a string of length m, there is an associated cost, that is calculated for possible change using array 'A' in the iteration (page 77, see code 'procedure select(A,C,S,Z)')

It would have been obvious to one of the ordinary skill in the art at the time of the invention to have modified Tang et al, Shanahan et al, Beeferman et al, and Hitachi's context model, to further include Srihari et al's method for implementing a Viterbi context model, which employs fringes. The combination would have allowed Tang et al's spell checking component to have "considered all alternatives for each of the m letters" (Srihari et al, page 75).

With regards to claim 34, which depends on claim 33, Tang et al, Shanahan et al, Beeferman et al, and Hitachi similarly teach a method comprising:

- *Employing, at least in part, a list of stop words to facilitate in determining at least one alternative in spelling;* as explained in claim 4, and is rejected under the same rationale.
- *Utilizing substring occurrence and co-occurrence statistics from at least one query log,* as explained in claim 1, and is rejected under the same rationale.
- *The query log comprising a histogram of queries asked over a time frame*, as explained in claim 9, and is rejected under the same rationale.

- *The substring occurrence and co-occurrence statistics from the query log are stored in a same searchable data structure*, as explained in claim 1, and is rejected under the same rationale.
- *Handling split substrings in the same manner as handling individual substrings*, as explained in claim 20, and is rejected under the same rationale.

With regards to claim 35, which depends on claim 34, Tang et al, Shanahan et al, Beeferman et al, and Hitachi similarly teach method comprising:

- *Changing at least one substring to another substring as an alternative spelling*, as explained in claim 23, and is rejected under the same rationale.
- *Halting the iterative correction process when all possible alternative spellings are less appropriate than a current set of alternative spellings*, as explained in claim 23, and is rejected under the same rationale.
- *The alternative spellings and their appropriateness are computed based on a probabilistic string distance and a statistical context model*, as explained in claim 24, and is rejected under the same rationale.

With regards to claim 36, which depends on claim 35, Tang et al, Shanahan et al, Beeferman et al, Hitachi, and Srihari et al similarly teach a method comprising:

- *Utilizing a searchable data structure extracted from at least one query log and at least one trusted lexicon to generate the set of alternative spellings for a substring*, as explained in claim 26, and is rejected under the same rationale.
- *Restricting the set of alternative spellings for each substring to within a probabilistic distance from an input substring, the restriction being imposed withn*

each iteration without limiting the iterative correction process as a whole, as explained in claim 27, and is rejected under the same rationale.

With regards to claim 41, Tang et al, Shanahan et al, Beeferman et al, and Hitachi similarly teach a device employing the method of claim 32, comprising at least one of a computer, a server, and a handheld device, as explained in claim 32, and is rejected under the same rationale.

7. Claim 25 is rejected under 35 U.S.C. 103(a) as being unpatentable over Tang et al (US Patent: 6,636,849 B1, issued: Oct. 21, 2003, filed: Nov. 22, 2000), Shanahan et al (US Application: US 2003/0033288 A1, published: Feb. 13, 2003, filed: Dec. 5, 2001), Beeferman et al (US Patent: 6,701,309 B1, issued: Mar. 2, 2004, filed: Apr. 21, 2000), and Hitachi (Derwent, published: Feb 16, 2001, Abstract), in further view of Brill et al (Microsoft Research: 'An Improved Error Model for Noisy Channel Spelling Correction', published: 2000, 8 pages).

With regards to claim 25, Tang et al, Shanahan et al, Beeferman et al, and Hitachi teach the *probabilistic string distance*, as similarly explained in the rejection for claim 24, and is rejected under the same rationale.

However, Tang et al, Shannahan et al, Beeferman et al, and Hitachi do not expressly the probabilistic string distance *comprises a modified context-dependent weighted Damerau-Levenshtein edit function that allows insertion, deletion, substitution, transposition, and long-distance movement of characters as point changes* (Brill et al, Page 2, First paragraph of section 2 'An Improved Error Model': whereas the Damerau-Levenshtein distance measures where the distance between two strings is the minimum

number of single character insertions, substitutions, and deletions, and transpositions: Thus, since the Damerau-Levenshtein distance does not require the edit distance to be only one, but instead, allows, the edit distance to be variable (but minimum), then long distance movement of characters is allowable)

Conclusion

8. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Wilson Tsui whose telephone number is (571)272-7596. The examiner can normally be reached on Monday - Friday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Stephen Hong can be reached on (571) 272-4124. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

WT 10/2/06

Wilson Tsui
Patent Examiner
Art Unit: 2178
October 2, 2006

STEPHEN HONG
SUPERVISORY PATENT EXAMINER